THE WELLCOME TRUST

Summary of the Report of the Second International Strategy Meeting on Human Genome Sequencing

Bermuda, 27th February - 2nd March 1997
Sponsored by the Wellcome Trust

Summary

- The principles enunciated at the first International Strategy meeting, of rapid data release and public access to the primary genomic sequence, were reaffirmed.
- Scientists and funding agencies should take the necessary steps to ensure that the principles are adhered to by all participating organisations.

Sequence Quality Standards

The following standards were agreed:

- The nucleotide error rate should be 1 error in 10,000 bases or less for most sequence.
- Assemblies should be verified by restriction digest using two or more restriction enzymes.
- Gaps in sequence. The agreed long term goal is no gaps, recognising that this is not yet routine.
- Closing gaps is the responsibility of the original sequencer.

The following proposals were endorsed by the participants:

- It was agreed that a useful trial to assess sequence accuracy would be to perform a data exchange exercise. Raw sequence data would be exchanged among sequencing centres, centres would reassemble the data and identify outright discrepancies or ambiguities with reference to the sequence submitted to the database. These would be resolved by further consultation or resequencing. The same data sets would be sent to two centres which would hopefully engender competition to detect errors.

- All sequence reads should be archived in a retrievable form.
- Sequencing centres should define explicitly how error rates and costs have been calculated.

Sequence Submission and Annotation
Sequence data should be classified simply as "finished" or "unfinished" and should be stored in distinct databases; consideration should be given to establishing a public database for unfinished sequence data.

Sequence annotation should be standardised if possible, and include the following information:

- Error estimation such as PHRED AND PHRAP data.
- Enzymes used to verify assemblies, and sizes of fragments produced.
- Exact details on how to assemble adjacent clones, with a minimum of 100 bp of overlapping (preferably unique) sequence between clones for verification.
- Gaps must be sized and the surrounding sequence oriented and ordered. The methods used for sizing, and reasons for not closing the gap should be stated.
- If features such as coding sequence and splice sites are included in the annotation, it should be stated if they were identified experimentally or by computer predictions.
- Unfinished sequence; it should be stated how near the sequence is to completion.

Potential development of a database listing all gaps in 'finished' sequence.

Sequence Claims and Etiquette
Mapping investment does not automatically entitle sequencing claims over the same region until a sequence ready map has been generated.

Potential conflicts with other sequencers to be resolved by early communication.

Collaborations with groups with a biological interest in a region should be subject to the same principles of data release and communication.

Investigate whether the Human Sequence Map Index should be relocated to be more closely associated with the other major human sequence databases.

Claims allowed on the Index:

- Duration - maximum 1 year.
- Size of region - minimum 1 Mb; regions to be defined by Généthon markers if possible, other agreed and available markers if not.
- Maximum amount - in the order of three times the sequence released by the centre in the preceding year.
- Sequence claims must span the entire region between, and including, the delimiting markers.

Next Meeting
To be held at the end of February 1998 in Bermuda (dates to be confirmed)

For further information regarding this summary, please contact:
Dr. Barbara Skene
The Wellcome Trust
183 Euston Road

London, UK, NW1 2BE
tel: (0)171 611 8888
fax: (0)171 611 8688
email: b.skene@wellcome.ac.uk

Return to HUGO Home Page

Last updated: 23 September 1998
by Marina Press